



TITLE:

多様体学習とデータ空間の幾何学 (統計的モデリングと予測理論の ための統合的数理研究)

AUTHOR(S):

小林, 景

CITATION:

小林, 景. 多様体学習とデータ空間の幾何学 (統計的モデリングと予測理論のための統合的数理研究). 数理解析研究所講究録 2017, 2057: 1-10

ISSUE DATE:

2017-10

URL:

<http://hdl.handle.net/2433/237186>

RIGHT:

多様体学習とデータ空間の幾何学

慶應義塾大学 JST さきがけ 小林 景

Kei Kobayashi

Keio University, JST PRESTO

1 概要

本稿では、機械学習やデータ解析の一分野である多様体学習についてのレビューを行った後、近年著者らが研究しているデータ空間の曲率に着目した距離の変換手法に関して紹介する。その後、データ空間の曲率という観点から、多様体学習と著者らの提案手法との違いを説明する。

2 多様体学習

統計学や機械学習などにおいて解析対象となるデータは、数ベクトルの形をしていることがほとんどである。よって自然にユークリッド空間上の点集合と同一視できる。しかし、実際にはデータはユークリッド空間の一部分（データ空間）にのみ存在する 경우가多く、そのことを仮定した上でデータ空間の情報をを用いた統計解析を行うことにより、解析精度を大きく向上させることができる。ここでいうデータ空間とは、データの分布のサポートを含むような、幾何学的対象（通常距離空間）のことである。ただし、場合によってはデータが近似的に（より高次元の空間に埋め込まれた）幾何学的対象の周辺に分布する場合にもデータ空間という用語を用いることがある。そういったデータ解析手法のうち、最も代表的なもののひとつが多様体学習である。本節では多様体学習について簡単にレビューを行う。なお、多様体学習の全般をまとめた文献としては、例えば [5] がある。

2.1 多様体学習の問題設定と目的

多様体学習の入力および出力は以下ようになる。

入力データ: $x_1, \dots, x_N \in \mathbb{R}^d$.

もしくは、距離行列 $D := (d_{ij})_{i,j=1}^N$. ただし, $d_{ij} := d(x_i, x_j)$.

出力データ: $z_1, \dots, z_N \in \mathbb{R}^{\tilde{d}}$.

ここで、出力次元 \tilde{d} (通常 $\tilde{d} < d$) は予め既知とする場合が多いが、クロスバリデーションやモデル選択などの統計的手法による最適な \tilde{d} を選ぶことも可能である。

データ点間の距離関係をなるべく保ちつつ、次元を減らすことが目的となるが、その応用上の動機は以下のことがあげられる。

- 視覚化（3次元以下に次元を落とすことによりデータを図示可能にする）
- 次元削減（データの圧縮、計算量削減が期待される）

- 特徴抽出（出力データの各座標をデータの特徴量とみなす）

これらは主成分分析の応用上の目的とほぼ同じである。また、特徴抽出は特徴抽出自体が目的である場合と、抽出された特徴を使つての統計解析やパターン認識が目的である場合がある。この後者の場合には、次元削減によって余計な次元によるオーバーフィッティングを回避するという効果も期待される。

なお、近年では多様体学習の多様化や他分野との融合にともない、より一般の「非線形次元削減」と同義で多様体学習という用語を用いる場合も多いが、本稿では、 k 近傍法などを通して局所的な距離構造に着目して次元削減するような、初期の多様体学習の手法を主に扱う。

2.2 古典的多次元尺度法

距離の変換を利用した次元削減の最も基本的な手法としては、古典的な多次元尺度法（Multi Dimensional Scaling, MDS）があげられる。古典的な多次元尺度法は以下のようなステップで行われる。

1. 距離行列 D の各要素を二乗: $D^{(o2)} = (d_{ij}^2)$,
2. 二重中心化: $B = -1/2JD^{(o2)}J$ ($J = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$),
3. 固有値分解: $B = P\Lambda P^\top$,
4. 低ランク近似: $B \simeq ZZ^\top$, $Z^\top = \sum_{k=1}^{\tilde{d}} \sqrt{\lambda_k} \mathbf{p}_k$.

ただし、 $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^N$, $\lambda_k = \Lambda_{kk}$ かつ \mathbf{p}_k^\top は P の第 k 行である。ここで、 $\tilde{X} := JX$ としたとき、 $J\mathbf{1} = \mathbf{0}$ かつ

$$\begin{aligned} B &= -\frac{1}{2}JD^{(o2)}J = -\frac{1}{2}J(\text{diag}(XX^\top)\mathbf{1}\mathbf{1}^\top - 2XX^\top + \mathbf{1}\mathbf{1}^\top\text{diag}(XX^\top))J^\top \\ &= JXX^\top J^\top = \tilde{X}\tilde{X}^\top \end{aligned}$$

となることを用いた。

古典的多次元尺度法は、その導出からもわかるように入力がユークリッド距離であることを仮定した手法である。また、線形射影による次元削減を行う。一方、多様体学習は一般に非線形写像により次元削減を行う。さらに、経験グラフ（データ点が頂点となるような距離グラフ）や局所的なデータ間距離を用いて構成した、より一般的な距離空間のデータや距離行列を入力とすることができる。多様体学習の種類は非常に多岐にわたるが、ここでは特に代表的な3つの手法に限って紹介する。

2.3 Isomap

Isomap は [7] において Tenenbaum, de Silva および Langford によって提案された以下のようなアルゴリズムである。

1. 各データ点の k 近傍 (k 番目までに近い点) を選び、各点への距離を長さとするような辺で結ぶ（グラフは連結になると仮定）,
2. 得られたグラフ (k 近傍グラフ) 上の最短経路により各データ点間の距離を計算する,

3. 得られた距離行列をもとに古典的な多次元尺度法を行う。

つまり、古典的な多次元尺度法との違いは、 k 近傍グラフの最短経路長を用いる点のみである。通常はもとのデータは数ベクトルであり、データ間距離はユークリッド距離を用いる。その場合は、 k 近傍部分は低次元ユークリッド距離で近似される。こういった局所的な距離をつなぎあわせて大域的な距離空間を作成するため、局所座標系を張り合わせて構成されるリーマン多様体の推定が可能となる。なお、 k の値が小さいほどデータ空間の局所的な構造をとらえるが、小さすぎるとデータのランダムネスの影響を受けやすく、また k 近傍グラフが連結にならない可能性が高くなる。

2.4 Locally Linear Embedding

Locally Linear Embedding は [6] において Roweis と Saul によって提案された以下のようなアルゴリズムである

1. 各点の k 近傍を選ぶ,
2. 各データ点 \mathbf{x}_i を k 近傍点 $kNN(\mathbf{x}_i)$ の線形結合で近似:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_i \|\mathbf{x}_i - \sum_{j: \mathbf{x}_j \in kNN(\mathbf{x}_i)} w_{ij} \mathbf{x}_j\|^2,$$

(なお、 $\arg \min_w f(w)$ は $f(w)$ を最小化する w の値を意味する.)

3. 重みベクトル \mathbf{w} を保ち、かつ低次元のベクトル集合を求める。

$$\arg \min_{\mathbf{z}_1, \dots, \mathbf{z}_N \in \mathbb{R}^d} \sum_i \|\mathbf{z}_i - \sum_{j: \mathbf{x}_j \in kNN(\mathbf{x}_i)} w_{ij} \mathbf{z}_j\|^2.$$

つまり、各データ点とその近傍点の線形結合関係をなるべく保つように、各データ点はより低次元のベクトルに置き換えられる。

2.5 Laplacian Eigenmap

Laplacian Eigenmap は [1] において Belkin と Niyogi によって提案された、以下のようなアルゴリズムである。

1. 各点の k 近傍を選ぶ,
2. 重み行列 W を以下のいずれかで定義する。
 - (1) 隣接行列
 - (2) 隣接行列にガウシアンを重みを付けたもの
(x_i, x_j が隣接しているならば、 $W_{ij} = \exp(-d(x_i, x_j)^2)$ そうでなければ $W_{ij} = 0$)
3. k 近傍グラフをもとに $N \times N$ ラプラシアン行列 $L := D - W$ を構成する。
ただし、 D は対角行列で $D_{ii} = \sum_{j=1}^N W_{ij}$.
4. ラプラシアン行列を用いた一般化固有方程式 $L\mathbf{v} = \lambda D\mathbf{v}$ を用いて低ランク近似する。

ここで、一般化固有方程式 $L\mathbf{v} = \lambda D\mathbf{v}$ は、 $ZDZ^T = I$ という制約条件のもとで、 $\sum_{i,j} W_{ij} \|z_i - z_j\|^2$ を最小化することに対応する。

また、 W が隣接行列のときは L はグラフラプラシアンと一致する。統計的な正則条件のもと、リーマン多様体の Laplace-Beltrami 作用素の固有ベクトルに収束することが示される。つまり、Laplacian Eigenmap は、グラフやリーマン多様体のラプラシアンをなるべく保ちつつ（大きな固有値に対応する固有空間を保つという意味で）次元削減を行っているとも考えられる。ラプラシアンを保つという発想は、データ点がある多様体の上を拡散運動によって動いている、もしくは定常状態にあると仮定できるような場合には、特に自然な方針であるといえるだろう。

2.6 計算量

最後に、上に述べた三種類の手法について計算量を考察する。 N を標本サイズ、 k を近傍数、 d を入力次元、 \tilde{d} を出力次元とすると、全ての手法に必要な k 近傍計算は、例えば BallTree アルゴリズムで $O[d \log(k) N \log(N)]$ である。また、Isomap で必要な最短経路計算は、Dijkstra 法もしくは Floyd 法で $O[N^2(k + \log(N))]$ である。Local Linear Embedding で必要な重み行列計算および、Laplacian Eigenmap で必要な Laplacian 計算は $O[dNk^3]$ である。さらに、全ての手法に必要な固有値分解の計算は $O[\tilde{d}N^2]$ である。

以上をまとめると、それぞれの手法の計算量は

$$\begin{aligned} \text{Isomap: } & O[d \log(k) N \log(N)] + O[N^2(k + \log(N))] + O[\tilde{d}N^2], \\ \text{Locally Linear Embedding: } & O[d \log(k) N \log(N)] + O[dNk^3] + O[\tilde{d}N^2], \\ \text{Laplacian Eigenmap: } & O[d \log(k) N \log(N)] + O[dNk^3] + O[\tilde{d}N^2]. \end{aligned}$$

通常は $k, d, \tilde{d} \ll N$ であるから、各手法ともに第二項もしくは第三項が主項となる。特に Isomap における最短経路計算は、標本サイズ N が大きいときはボトルネックとなる。

3 データ空間の曲率に着目した距離変換法

著者は近年 Henry P. Wynn (London School of Economics) とともにデータ空間の距離をその曲率に着目して変換して、データ解析に応用する手法とを提案している [4]。本節ではその内容を短くまとめて説明する。我々が提案する距離変換手法はその目的において、多様体学習と共通する点と本質的に異なる点があるが、それらについては次節にまとめる。

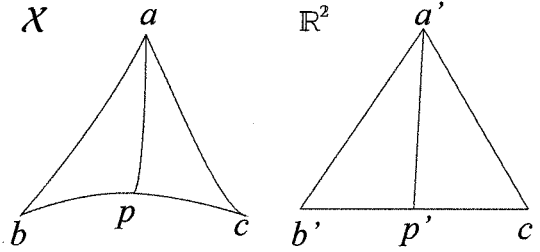
3.1 提案手法の目的

通常の統計解析においては、データが線形空間内に分布すると仮定し、線形性を利用した手法を用いる。またその理論評価も線形性に大きく依存する。近年では関数空間などの線形位相空間上のデータ解析手法の研究が盛んであるが、線形性に依存するという点では古典的なユークリッド空間上のデータ解析と共通している。一方、リーマン多様体など線形性をもたない距離空間上のデータ解析についてもいくつか既存研究がある。先にあげた多様体学習もその一例である。

本研究では、多様体学習の目的である距離の近似による次元削減とは異なり、距離関数をパラメータを導入して変換し、距離関数の集合（クラス）を構成する．このように距離関数のクラスを構成しておけば、データに依存してクロスバリデーション等の手法でその中から適当なものを選択することができる．一方、通常データ解析や機械学習では、損失関数のクラスの中から適当なものを選ぶ．それと比べると、距離関数のクラスを用いることにより、より積極的に幾何学を導入した手法や理論の構築が可能となる．

3.2 データ空間の曲率

本研究では、データ空間は測地距離空間であると仮定して、その曲率を $\text{CAT}(k)$ 特性という基準を用いて評価する．ここで測地距離空間とは、任意の二点間の距離が、それを結ぶ最短の曲線長と一致するような距離空間のことである．また、測地距離空間 (\mathcal{X}, d) が $\text{CAT}(0)$ 空間であるとは、任意の $a, b, c \in \mathcal{X}$ と $a', b', c' \in \mathbb{R}^2$ が $\|a' - b'\| = d(a, b), \|b' - c'\| = d(b, c), \|c' - a'\| = d(c, a)$ 等をみたすとき、測地線 \tilde{bc} 上の点 p と、 $d(b, p) = \|b' - p'\|$ をみたす辺 $\tilde{b'c'}$ 上の点 p' に対して、 $d(a, p) \leq \|a' - p'\|$ が成り立つことである．



直感的には、 \mathcal{X} 上の測地三角形がユークリッド空間上の対応する三角形より「へこむ」ような空間であり、局所的には非正曲率を持つ．

より一般に $k \in \mathbb{R}$ については、 $\text{CAT}(0)$ 空間と同様に $\text{CAT}(k)$ 空間が定義できる．ただし、3 辺の長さの和が $2\pi/\sqrt{\max(k, 0)}$ 以下であるような測地三角形の場合に限り、また、ユークリッド平面の代わりに、(リーマン断面曲率の意味で) 曲率 k の定曲率曲面上の測地三角形と比較する．この定義によると、リーマン多様体の場合には $\text{CAT}(k)$ 空間ならば局所的には(断面曲率の意味で) 曲率が k 以下であることがいえる．これより、測地距離空間が $\text{CAT}(k)$ 空間であるとき、 $k' \geq k$ ならば $\text{CAT}(k')$ 空間でもあることがわかる．

測地距離空間 (\mathcal{X}, d) 上の標本 x_1, \dots, x_n の内測平均 (intrinsic mean, Fréchet mean) は以下で定義され、その一意性は \mathcal{X} の $\text{CAT}(k)$ 特性に依存することが知られている．

$$\hat{\mu}_d = \arg \min_{m \in \mathcal{X}} \sum_{i=1}^n d(x_i, m)^2.$$

また、対応する分散を以下で定義できる．

$$\hat{\text{Var}}_d = \min_{m \in \mathcal{X}} \frac{1}{n} \sum_{i=1}^n d(x_i, m)^2.$$

さらに、関数 $f(m) = \sum_{i=1}^n d(x_i, m)^2$ は Fréchet 関数とよばれ、測地距離空間上のデータの特徴を表す最も基本的な関数の一つである．ここでは標本についての平均をとったが、母集

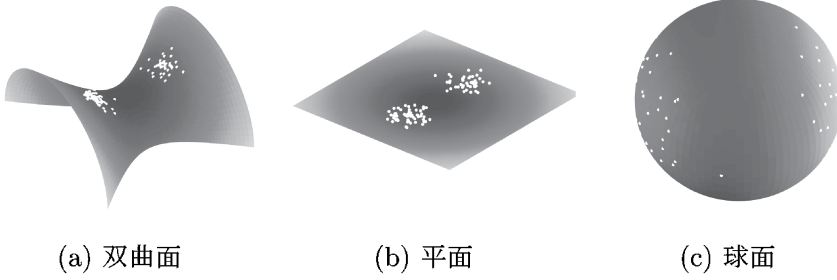


図 1: それぞれ (a) 双曲面 (定曲率 $c = -1$), (b) 平面 ($c = 0$), (c) 球面 ($c = 1$) 上のデータに対する Fréchet 関数 f の値. (濃い色が小さな値)

団分布についての期待値で定義すると確率分布についての内測平均, 対応する分散, Fréchet 関数も定義できる.

図 1 は, (a) 双曲面 (定曲率 -1), (b) 平面 (定曲率 0), (c) 球面 (定曲率 1) 上のデータに対して, Fréchet 関数の値を図示したものである. 正の曲率を持つ球面のみが, $f(\mu)$ の極小値を複数持つことがわかる. 曲率と内測平均の一意性の関係についての詳細は, 例えば [2] を参照.

次に, 以下で導入する β 距離を用いた内測平均と関連する事項として, 外測平均を説明する. 測地距離空間 (\mathcal{X}, d) が他の距離空間 $(\tilde{\mathcal{X}}, \tilde{d})$ に埋め込まれているとき, 埋め込まれた空間の距離を用いて

$$\hat{\mu} = \arg \min_{m \in \mathcal{X}} \sum_{i=1}^n \tilde{d}(x_i, m)^2.$$

としたものを外測平均 (extrinsic mean) とよぶ. \min は \mathcal{X} 上でとることに注意. 埋め込む空間 $(\tilde{\mathcal{X}}, \tilde{d})$ は通常ユークリッド空間で, 測地距離を用いた内測平均より計算が簡単であるために応用上有効である.

本提案手法では, 上記の Fréchet 関数をパラメータ $\alpha \in \mathbb{R}, \beta \in (0, \infty], \gamma \geq 1$ を導入して以下のように一般化する.

$$f_{\alpha, \beta, \gamma}(m) = \sum_i g_{\beta}(d_{\alpha}(x_i, m))^{\gamma}$$

ただし, $d_{\alpha, g_{\beta}}$ については以下で説明する. これにより, 内測平均および分散も一般化される. また, 一般化された Fréchet 関数は非凸となり得るので, その極小点はクラスタリングにおけるクラスター中心の候補となる.

3.3 α 距離変換

リーマン多様体のデータ空間 \mathcal{M} 上の確率分布の密度関数 f に対して, 二点 $x_0 = z(0)$, $x_1 = z(1)$ を結ぶ曲線 $\Gamma = \{z(t), t \in [0, 1]\}$ の長さを $d_{\Gamma, \alpha}(x_0, x_1) = \int_0^1 s(t) f^{\alpha}(z(t)) dt$ で定義する. ただし, $s(t)$ は曲線に沿った微小距離単位である. この曲線長を用いた測地距離を α 距離とよび, もとのデータ空間の測地距離の α 距離変換とよぶ. また, リーマン多様体に限らず, 一般の可測な測地距離空間に対しても同様に α 距離変換を定義可能である. $\alpha = 0$ は

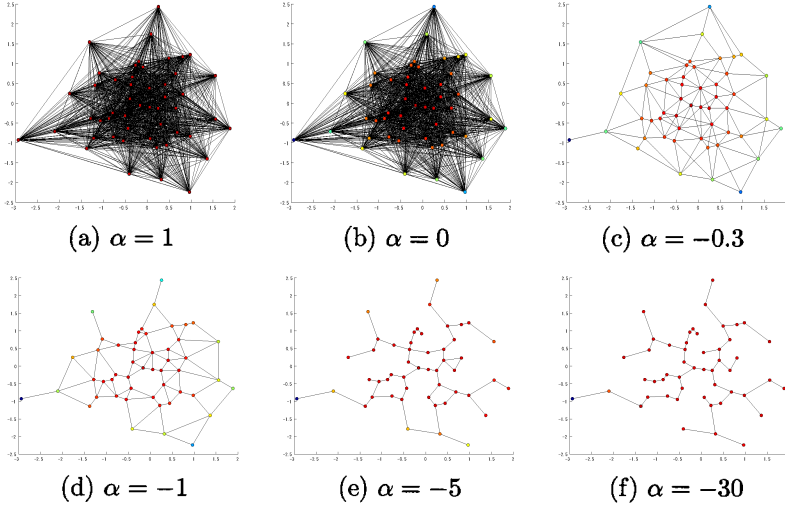


図 2: 二次元正規分布からの標本に対して, α の値を変化させたときの経験グラフの変化. 各データ点の位置は固定し, 辺の有無のみ変化させて表示した.

元のデータ空間の距離に対応する. また, 1 次元のときは $\alpha = 1$ の α 距離を用いた内測平均はメディアンと一致する.

しかし, $d_{T,\alpha}$ の計算は密度関数の推定および積分計算を要し, 実データ解析で用いることは一般に困難である. そこで, α 距離の経験グラフ (データ点を頂点とする距離グラフ¹) を用いた離散近似版を以下のように計算する.

1. 標本を頂点とするグラフを構成する (完全グラフ, k -NN グラフ等).
2. 各辺の長さ d_{ij} を $d_{ij}^{1-\alpha}$ に変換する.
3. グラフでの最短経路長で標本間の距離を定義する.

この離散版の α 距離およびそれを用いた Fréchet 平均が α の値にどのように依存するかを調べるために, 測地部分グラフを用いる. 距離グラフの測地部分グラフ (geodesic sub-graph) とは, どの頂点対を結ぶ最短経路にも含まれない辺を除いた部分グラフである. データ点を頂点とする距離グラフにおいて, Fréchet 関数の測地部分グラフ上での値は (よって全頂点上の値も), 測地部分グラフのみによって定まる.

このように定義された測地部分グラフは, α の減少にともない縮小する. 図 2 はその一例である. さらに, α の減少にともない $CAT(k)$ となる領域が増大すること (曲率の減少に対応), および極限が最小全張木になることなどを [4] において証明した. つまり, データ空間の測地距離の経験グラフ近似に対し, α の値を変化させることにより, $CAT(k)$ 特性の意味での曲率を単調に変化させることが可能である.

¹距離グラフとは, ここでは各辺 $e \in E$ に長さ (非負の重さ) が定義されている連結グラフ $G = (V, E)$ のことを指すとする.

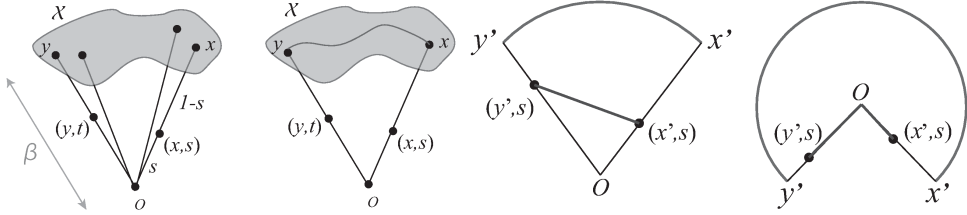


図 3: 計量錐のイメージ (測地線の張る錐を扇状に開いて平面に埋め込む)

3.4 β 距離変換

α 距離変換は測地距離を測地距離に変換する. 一方, 以下で説明する β 距離は必ずしも測地距離になるとは限らない. よって, β 距離では $\text{CAT}(k)$ 特性は定義できず, どのような意味で曲率に着目した距離変換を行うかに工夫が必要となる.

パラメータ $\beta \in (0, \infty]$ による距離 d の変換を以下で定義する.

$$d_\beta(x_0, x_1) = g_\beta(d(x_0, x_1)) \quad (\beta > 0),$$

ただし

$$g_\beta(z) = \begin{cases} \sin(\frac{\pi z}{2\beta}), & \text{for } 0 \leq z \leq \beta, \\ 1, & \text{for } z > \beta. \end{cases}$$

この距離変換を β 距離変換とよび, d_β を β 距離とよぶ.

β 距離空間は, 一般に測地距離空間ではないので $\text{CAT}(k)$ 特性を定義できない. 一方, β 距離による内測平均は, 動径 β の計量錐に埋め込まれた空間の外測平均として解釈できる.

ここで, 測地距離空間 \mathcal{X} から構成される計量錐 (metric cone) $\tilde{\mathcal{X}}$ とは, 集合 $\mathcal{X} \times [0, 1]/\mathcal{X} \times \{0\}$ 上の各点对 (x, s) と (y, t) に以下の距離を導入したものである:

$$\tilde{d}((x, s), (y, t)) = \sqrt{t^2 + s^2 - 2ts \cos(\pi \min(d_{\mathcal{X}}(x, y), 1))}.$$

$\tilde{\mathcal{X}}$ は測地距離空間になることが証明できる (図 3 を参照). 直観的には, 「原点」 O と測地距離空間 \mathcal{X} の各点とを結ぶ長さ 1 の「線分」の集合に, うまく測地距離を定義したものと考えられる. ユークリッド空間に埋め込めるとは限らない. 以下のように距離を定義すると, 動径 (「線分」の長さ) を r とした計量錐を構成することもできる:

$$\tilde{d}_r((x, s), (y, t)) = \sqrt{t^2 + s^2 - 2ts \cos(\pi \min(d_{\mathcal{X}}(x, y)/r, 1))}.$$

ここで, β が小さくなると, $\text{CAT}(k)$ 特性の意味で計量錐の曲率が小さくなることが証明できる ([4] 参照). つまり, β を調整することにより, データ空間の埋め込まれている計量錐の曲率を単調に変化させることができる.

β 距離変換を導入する動機として, Fréchet 関数を非凸にして, 複数個の極小点 (Karcher 平均とよばれる) をもたせることがある. (ユークリッド空間では距離関数は凸関数であるため, Fréchet 関数も凸関数となり常に唯一の最小値 (内測平均) をもつ. これにより, Fréchet 関数の極小点をクラスター中心の候補に用いることができ, クラスタリングに応用できる. つまり, 埋め込まれた錐の曲率を変化させることにより, クラスタ中心の候補数を調整できる. 本研究はおそらく, ユークリッド空間以外に埋め込まれたデータ空間に対する外測平均をデータ解析に応用した最初の例である.

4 提案手法と多様体学習の比較 — 曲率の観点から —

既存の多様体学習の手法と、新提案手法を順に説明した。これらの間には共通する部分と、本質的に異なる部分がある。本節では、両手法の共通点と差異を説明し、また多様体学習の既存研究から示唆される新提案手法の問題点と今後の発展性についてまとめる。

まず両手法の共通点としては、線形位相空間ではなく、多様体を含む測地距離空間をデータ空間とするようなデータ解析手法であることがあげられる。古典的な統計解析は、線形空間上のデータを線形的な手法で解析するので、測地距離空間上のデータ解析に発展させるためには工夫が必要となる。その一つが、経験グラフを用いてデータ間距離を計算するという点であり、この点も共通する。特に、Isomap はグラフ上での最短経路長を用いるという点で新提案手法と近く、アルゴリズムも共に Dijkstra 法や Floyd 法を用いる。それゆえ、この部分に計算量のボトルネックがあるというのが共通の課題であるといえる。

一方、両手法の差異として、多様体学習は次元削減が目的であるのに対し、新提案手法では距離の候補の作成が目的であるという大きな違いがあげられる。多様体学習では、予測問題のように最終的に「真の」データ空間が必要でない場合であっても、まず「真の」データ空間を推定する。他方、新提案手法のように距離の候補を作り、予測問題のために適当な距離を選択するという手法は、「真の」データ空間の推定の部分は省略している。これは、ベイズ予測や機械学習における、「真の」パラメータの推定やモデルの同定を省略して予測に最適な事前分布を選択するという発想を、より推し進めたものであるといえる。

次に、提案手法では主役であった曲率に着目して、多様体学習の3手法を再考察する。まず、Isomap と Locally Linear Embedding は、局所距離を低次元ユークリッド空間（曲率 0）で近似している。大雑把に言えば、データ空間の多様体の（ほぼゼロであると仮定された）曲率を近似するような低次元空間を求めているということができよう。Laplace Eigenmap についても以下のように考えると、曲率を近似するアルゴリズムであるということができる。

まず、リーマン多様体のラプラシアン（Laplace-Beltrami 作用素）について、以下のようなリーマン曲率との関係が知られている。例えば [3] を参照。

1. コンパクトな n 次元リーマン多様体 (M, g) について、ある定数 $\kappa > 0$ が存在しリッチ曲率が $Ric(X, X) \leq \kappa g(X, X)$ を満たすとする。このとき、Laplace-Beltrami 作用素の 0 の次に小さな固有値 $\lambda_1 > 0$ に対して、 $\lambda_1 \geq \frac{n}{n-1}\kappa$ が成り立つ。
2. さらに等号が成り立つのは、半径 $1/\sqrt{\kappa}$ の n 次元球面 $S^n(1/\sqrt{\kappa})$ と等距離なものに限る。

よって、Laplace-Beltrami 作用素は曲率の情報を持ち、その意味では Laplace Eigenmap はデータ空間の曲率の情報をなるべく保つような次元削減であるということができる。

また、多様体学習で計算された距離を、新提案手法を用いて変換することも可能である。新提案手法では、データの空間の次元は低い方がより信頼性が高い解析結果になることが期待されるため、多様体学習によりデータ空間の曲率をなるべく保ちつつ次元削減するという前処理は合理的であるといえる。

ところで、提案手法の曲率は、グラフ上の CAT(k) 特性を用いて議論された。グラフに関するこれ以外の曲率の定義や、測地距離空間や多様体上の（経験分布による離散近似ではなく連続的なデータ空間に対する）距離の変換とその曲率の関係の解析は課題である。さら

に、経験グラフによる測地距離近似精度の理論評価についても今後の課題であり、提案手法のみならず多様体学習についても重要な理論研究テーマといえるだろう。

参考文献

- [1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, volume 14, pages 585–591, 2001.
- [2] Abhishek Bhattacharya and Rabi Bhattacharya. *Nonparametric inference on manifolds: with applications to shape spaces*. Number 2. Cambridge University Press, 2012.
- [3] Martin Vito Cruz. The spectrum of the laplacian in riemannian geometry. *preprint*, 2003.
- [4] Kei Kobayashi and Henry P Wynn. Empirical geodesic graphs and cat (k) metrics for data analysis. *arXiv preprint arXiv:1401.3020*, 2014.
- [5] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007.
- [6] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [7] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

223-8522 横浜市港北区日吉 3-14-1 慶應義塾大学 理工学部
E-mail: kei@math.keio.ac.jp

慶應義塾大学 小林 景